大粒香水稻叶绿体基因组特征分析

吴朝昕，刘雪薇*，李祖军，龙武华，宫彦龙，朱速松
（贵州省农业科学院水稻研究所，贵阳，550006）

摘要：大粒香作为贵州重要的优质稻资源，推广种植面积大，且在乡村振兴过程中为社会带来了较高的经济效益，但目前对大粒香基因组学理论研究报道不多。为了揭示大粒香水稻叶绿体基因组特征及系统发育关系，该文对大粒香叶绿体进行测序，分析其基因组特征。结果表明：（1）大粒香水稻 cpDNA 全长 134 563 bp，包括一个 LSC（80 864 bp），一个 SSC（12 347 bp）和两个 IRs（20 676 bp）。（2）注释到 129 个基因，可分为蛋白编码、tRNA 和 rRNA 三类基因，数量分别为 85 个、36 个和 8 个。（3）密码子偏好分析显示，大粒香 cpDNA 密码子偏好 A 碱基或者 U（T）碱基，亮氨酸密码子使用了 1 944 次，最多；半胱氨酸的密码子仅使用 198 次，最少。（4）检测到 129 个 SSR，其中有 95 个单核苷酸重复，且大部分 SSR 序列由 A/T 碱基组成。（5）系统发育分析结果表明，大粒香与热带粳稻亲缘关系最近聚为一类。该研究揭示了大粒香叶绿体基因组信息，并确定了大粒香系统发育所属分支。

关键词：水稻，大粒香，叶绿体基因，SSR，系统发育分析

Analysis of chloroplast genome of rice Dalixiang

WU Chaoxin, LIU Xuewei*, LI Zujun, LONG Wuhua, GONG Yanlong, ZHU Susong
（*Guizhou Rice Research Institute, Guizhou Academy of Agricultural Sciences*,
550006, Guiyang, China）

Abstract: As an important high-quality rice resource in Guizhou province, Dalixiang has a large planting area, and has brought higher economic benefits to society in the process of rural revitalization. However, there are few theoretical researches on the genomics of Dalixiang. In order to reveal the characteristics and phylogenetic relationships of chloroplast genome of large grain perfume rice, the chloroplasts of grancas fragrances were sequenced and their genomic characteristics were analyzed. The results were as follows: (1) The chloroplast genome of Dalixiang was 134 563 bp, including LSC(80 864 bp),SSC(12 347 bp)and two IRs(20 676 bp). (2) There were 129 genes annotated in the chloroplast genome of Dalixiang, which could be divided into protein coding, tRNA and rRNA, with 85, 36 and 8 genes respectively. (3) Codon bias

基金项目：黔农科院青年基金([2018]101 号)；贵州省科技基础项目(黔科合基础[2020]1Y101);黔农科院引导项目[2018]03 号；黔科合支撑项目([2019]2304 号)；黔科中引地项目([2019]4001)；黔科合平台人才项目([2017]5719)；黔科合基础（ZK[2021]124）[Supported by Youth Fund of Guizhou Academy of Agricultural Sciences([2018]101); Project of Guizhou Science and Technology (QianKeHeJiChu[2020]1Y101); Project of Guizhou Academy of Agricultural Sciences([2018]03); Project of Guizhou Science and Technology Cooperation[2019]2304; Central Guided by Guizhou Science and Technology Cooperation([2019]4001) ; Platform Talent of Guizhou Science and Technology Cooperation([2017]5719); Project of Guizhou Science and Technology (QianKeHeJiChu-ZK[2021]124)]。

第一作者：吴朝昕(1995-)，硕士，主要从事水稻分子育种，(E-mail) wuchaoxin1995@163.com。
***通信作者：**刘雪薇，硕士，主要从事水稻分子育种，(E-mail) 627605375@qq.com。

analysis of Dalixiang showed that leucine was most frequently used and that cysteine was used least frequently, and most codons ended in A/U(T). (4) The total number of SSR loci in the cpDNA of Dalixiang was 129, ninety-five of which are mononucleotide and most of SSR was composed of nucleobase A/T. (5) Phylogenetic analysis shows an affiliation relationship between Dalixiang and Tropical Japonica. This study reveals the characteristic information of Dalixiang chloroplast genome, and identifies the phylogenetic status of Dalixiang.
Key word: rice,dalixiang, chloroplast genome, SSR, phyletic evolution

叶绿体是植物获取能量的重要细胞器，有着独立于核基因组的遗传体系，其基因序列保守,基因结构重排事件远低于核基因组,结构简单,一般为母系遗传,因此被用来揭示物种的进化与亲缘关系（李绪英等，2011；Li et al., 2018; Zhang et al., 2018; Jeon & Kim., 2019;朱斌等，2021）。由于测序科技的发达，不少生物实现了叶绿体基因组测序，NCBI 中收录的叶绿体基因逐渐增多，利用叶绿体基因研究亲缘关系的报道也不断增多。在水稻中，Fan 等（2020）利用 cpDNA 分析了 33 个稻属物种的亲缘关系，研究显示 *O. sativa* voucherHSAGSDYD1802 与 *O. sativa* cultivar TN1、*O. sativa* cultivar RP Bio-226 和 *O. sativa* cultivar IR8 的亲缘关系最近。Fang 等（2017）通过分析 cpDNA 探讨了根茎野生稻与其他 13 个稻属物种的亲缘关系，研究显示与根茎野生稻关系较近的是药用野生稻，且属于 CC 基因类型。在其他物种中，张慧等（2018）利用 cpDNA 分析了益母草及其他 16 个物种的亲缘关系，研究结果很好的解决了野芝麻亚科的进化关系。郑祎等（2020）用大花君子兰叶绿体基因序列与 10 个百合科、5 个兰科、4 个鸢尾科及 5 个石蒜科共 24 个物种的叶绿体基因组序列进行系统发育分析，研究结果支持大花君子兰属于石蒜科，并且使用其中 23 个物种叶绿体基因组中 *ycf*2 进行亲缘关系分析，发现叶绿体基因组中 *ycf*2 可以代替叶绿体基因组全长进行亲缘关系分析。

大粒香是贵州省水稻研究所选育，具有稻米粒大且香特点，并且大粒香在乡村振兴过程中为社会带来了较高的经济效益（蒋志谦，2008；罗仁发等，2012）。目前，对大粒香基因组学、品质形成等理论研究的文献报道不多。基于大粒香在贵州优质稻发展过程中的重要性，为进一步从基因组水平认识和改良大粒香，本研究，以大粒香 DNA 为材料进行测序，构建大粒香 cpDNA 图谱，分析密码子的使用，重复序列，分析亲缘关系，拟揭示大粒香叶绿体基因组以下信息：（1）大粒香 cpDNA 的基本特征大小。（2）大粒香叶绿体基因密码子偏好情况。（3）大粒香叶绿体基因组系统发育所属分支。

1 材料与方法

1.1 试验材料

以“大粒香”为试验材料，2019 年冬季种植于海南三亚师部农场基地，2020 年 2 月选取无病虫害、长势良好的水稻叶片，冲洗、擦干液氮速冻暂存，后用干冰保存寄回贵州，置于-80℃保存，用于提取大粒香 DNA。

1.2 试验方法

1.2.1 大粒香 DNA 提取及测序

使用 TIANGEN 植物 DNA 试剂盒提取大粒香总 DNA，测序公司检测合格后，构建文库，进行测序。

1.2.2 叶绿体基因组序列组装与注释

将原始数据,除去有污染、低质量的片段，使用 SPAdes 软件拼接，然后组装。在使用 CPGAVAS2 进行基因注释，最后利用 OGDRAW 将大粒香的叶绿体基因图谱绘制出。

1.2.3 密码子使用分析

使用 CodonW 进行密码子使用分析。

1.2.4 重复序列分析

使用 vmatch 完成大粒香 cpDNA 的长重复序列的查找。大粒香 cpDNA 的 SSR 筛选则使用 MISA 软件，该软件的检测参数为，单核苷酸大于 8 时被检测，2 核苷酸和 3 核苷酸大于 4 时被检测。

1.2.5 系统进化分析

为探究大粒香与其他稻属物种的亲缘关系，从 NCBI 中下载了 12 个稻属物种和 2 种禾本科近源物种的叶绿体基因组（绿竹和高粱），使用 RaxML 软件构建系统发育树。

2 结果与分析

2.1 大粒香叶绿体基因组序列特征

大粒香 cpDNA 全长为 134 563 bp，分为 LSC（80 864 bp），GC 含量为 37.09 %、SSC（12 347 bp），GC 含量为 33.37 %和 IRs 区（20 676 bp），GC 含量为 44.41 %。在大粒香 cpDNA 中注释 129 个基因（表 1），可分为三类：蛋白编码基因、tRNA 基因和 rRNA 基因，其数量分别为 85、36 和 8。其基因功能主要为与自身复制能力有关、与光合作用有关、其他关系和未知功能有关等四种。在蛋白编码基因中，*rps* 基因的数量最多有 16 个；而 *cem*A、*inf*A、*ccs*A、*rbc*L、*mat*K、*acc*D、*clp*P 等基因数量仅有 1 个。其中有二十个基因出现在 IR 重复区内，分别为 *ndh*B、*ycf*1、*rps*12、*rps*7、*rps*15、*rps*9、*rpl*2、*rpl*23 等八个蛋白编码基因；*rrn*23S、*rrn*4.5S、*rrn*16S、*rrn*5S 等 4 个核糖体 RNA；及 *trn*N-GUU、*trn*H-GUG、*trn*A-UGC、*trn*L-CAA、*trn*T-CGU、*trn*V-GAC、*trn*R-ACG、trnM-CAU 等 8 个转运 RNA（表 1）

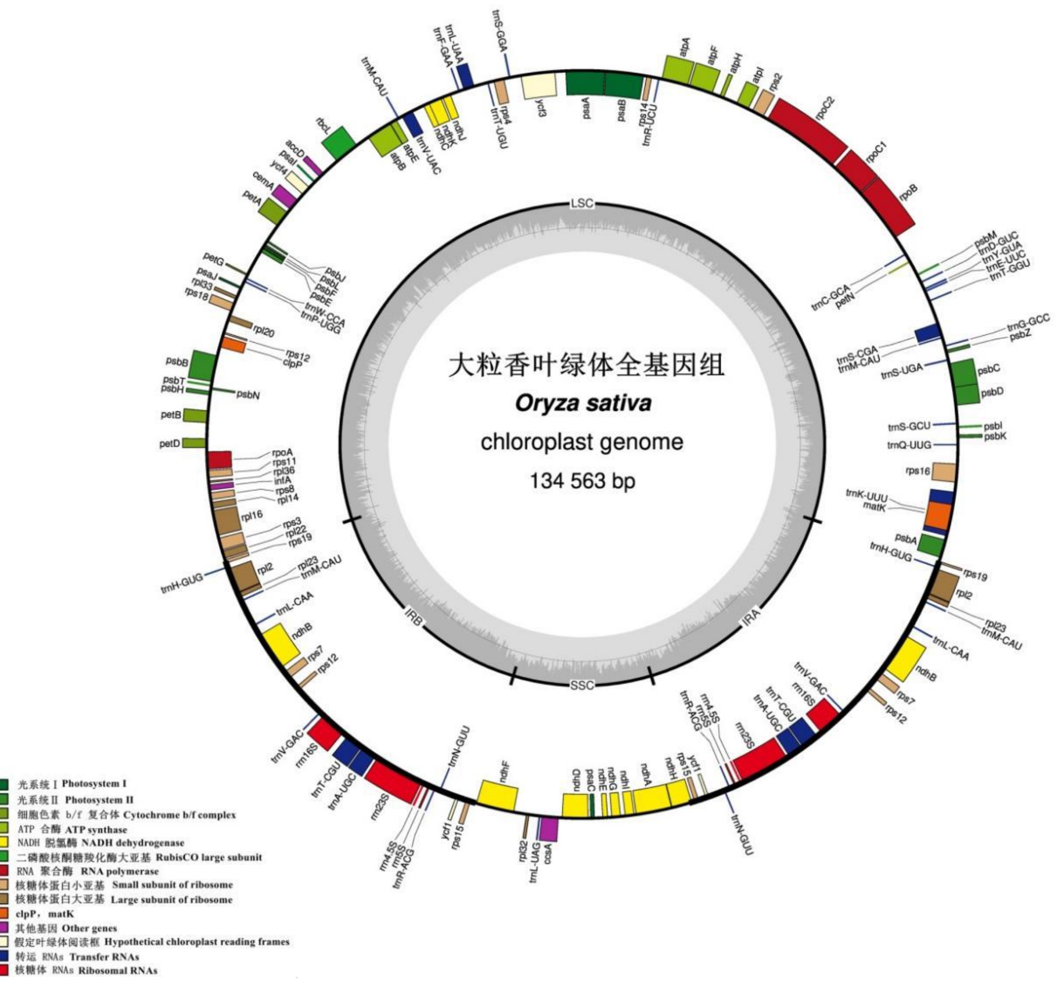


图 1 大粒香叶绿体基因组图谱

Fig. 1 Gene map ofDalixiang chloroplast genome

表 1 大粒香叶绿体基因组注释基因列表

Table 1 List of genes found in Dalixiang chloroplast genome

功能 Function	基因类别 Gene category	基因名称 Gene Name
光合作用 Photosynthesis	光系统Ⅰ Photosystem Ⅰ	*psa*C,*psa*I, *psa*B,*psa*J, *psa*A
	光系统Ⅱ Photosystem Ⅱ	*psb*L, *psb*D, *psb*N, *psb*B, *psb*T, *psb*H, *psb*E, *psb*A, *psb*I, *psb*K, *psb*J, *psb*C,*psb*M,*psb*Z, *psb*F
	细胞色素 b/f 复合体 Cytochromeb/fcomplex	*pet*A, *pet*B, *pet*D, *pet*N, *pet*G
	ATP 合酶 ATPsynthase	*atp*I, *atp*E ,*atp*A ,*atp*H, *atp*B, *atp*F
	NADH 脱氢酶 NADH dehydrogenase	*ndh*B*, *ndh*D, *ndh*E, *ndh*I, *ndh*F, *ndh*H, *ndh*G, *ndh*C, *ndh*K, *ndh*A, *ndh*J
	核酮糖-1,5-二磷酸羧化酶大亚基 RubisCO large subunit	*rbc*L
自我复制 Self replication	RNA 聚合酶 RNA polymerase	*rpo*C1, *rpo*C2, *rpo*A, *rpo*B
	核糖蛋白小亚基(SSU)	*rps*2, *rps*3, *rps*16, *rps*11, *rps*4, *rps*19*, *rps*18, *rps*14, *rps*8, *rps7**, *rps*15*, *rps*12*
	核糖蛋白大亚基(LSU)	*rpl*20, *rpl*32, *rpl*36, *rpl*22, *rpl*23*, *rpl*16, *rpl2**, *rpl*14, *rpl*33
	转运 RNAs TransferRNAs	*trn*Q-UUG,*trn*V-UAC, *trn*N-GUU*, *trn*C-GCA, *trn*H-GUG*,*trn*T-UGU,*trn*A-UGC*,*trn*R-UCU,*trn*L-CAA*, *trn*P-UGG, *trn*L-UAG, *trn*K-UUU,*trn*W-CCA, *trn*G-GCC, *trnT*-CGU*, *trn*V-GAC*, *trn*L-UAA, *trn*T-GGU, *trn*S-GCU, *trn*D-GUC, *trn*E-UUC, *trnS*-UGA, *trn*R-ACG*, *trn*F-GAA,*trn*Y-GUA, *trn*S-GGA, *trn*S-CGA,*trn*M-CAU*
	核糖体 RNAs Ribosomal RNAs	*rrn*23S*, *rrn*4.5S*, *rrn*16S*, *rrn*5S*
其他基因 Other genes	成熟酶 matKgene	*mat*K
	起始因子 initiation factor	*inf*A
	合成 C 型细胞色素的基因 c-type cytochrom synthesis gene	*ccs*A
	乙酰辅酶 A 羧化酶 Subunit of Acetyl-CoA-carboxylase	*acc*D
	被膜蛋白 Envelop membrane protein	*cem*A
	蛋白酶 Protease	*clp*P
未知功能 Unknown function	保守假定阅读框 Conserved hypothetical reading frame	*ycf*1*, *ycf*3, *ycf*4

注：*表示该基因位于反向重复区内。

Note: * indicates the gene is located in the inverted repeats.

统计结果显示，大粒香 cpDNA 中有内含子的基因共十七个，其中 *ycf*3 有两个，剩余十六个基因只有一个。其次,*trn*K-UUU 的内含子碱基数最多，而 *trn*L-UAA 的内含子碱基数最少（表 2）。

表 2 大粒香 cpDNA 含有内含子的基因

Table 2 Genes with introns in the Dalixiang chloroplast genome

名称 Name	外显子 I Exon I	内含子 I Intron I	外显子 II Exon II	内含子 II Intron II	外显子 III Exon III
*rpl*2(x2)	393	661	427		
*rps*16	212	811	44		
*rpl*16	401	1 060	8		
*trn*T-CGU(x2)	31	929	58		
*trn*A-UGC(x2)	36	812	35		
*trn*S-CGA	61	662	30		
*trn*K-UUU	35	2 488	37		
*trn*V-UAC	52	929	31		
*trn*L-UAA	35	540	50		
*ndh*B(x2)	776	710	758		
*ndh*A	541	986	549		
*atp*F	159	814	406		
*ycf*3	155	728	229	743	123

2.2 大粒香叶绿体基因组密码子使用分析

在大粒香 cpDNA 密码子中，亮氨酸密码子使用了 1 944 次，最多；半胱氨酸的密码子仅使用 198 次，最少。其次，在编码大粒香 cpDNA 的密码子中有 30 个密码子偏好性>1，其中以 A 结尾的有 12 个，以 U（T）结尾的有 16 个，这表明大粒香 cpDNA 密码子偏好 A/U（T）碱基，这种情况也常常出现在杜梨、益母草等多种高等植物中（张慧等，2018；李泳潭等，2020；郑祎等，2020）（表 3）。

表 3 大粒香 cpDNA 密码子使用

Table 3 Codon Usage in chloroplast genome of Dalixiang

名称 Name	密码子 Codon	次数 Number	密码子偏好性 RSCU	名称 Name	密码子 Codon	次数 Number	密码子偏好性 RSCU
异亮氨酸	AUU	730	1.51	缬氨酸	GUU	400	1.56
Isoleucine	AUC	286	0.59	Valine	GUC	115	0.45
	AUA	435	0.9		GUA	376	1.46
苏氨酸	ACU	402	1.71		GUG	136	0.53
Threonine	ACC	179	0.76	丙氨酸	GCU	489	1.74
	ACA	249	1.06	Alanine	GCC	162	0.58
	ACG	108	0.46		GCA	331	1.18
赖氨酸	AAA	638	1.47		GCG	142	0.51

lysine	AAG	231	0.53	天冬氨酸 Aspartic acid	GAU	508	1.56
天冬酰胺 Asparagine	AAU	515	1.45	天冬氨酸 Aspartic acid	GAC	143	0.44
	AAC	193	0.55	谷氨酸	GAA	697	1.47
丝氨酸	UCU	331	1.56		GAG	249	0.53
Serine	UCC	254	1.2	甘氨酸	GGU	415	1.26
	UCA	223	1.05	Glycine	GGC	138	0.42
	UCG	110	0.52		GGA	514	1.56
	AGU	266	1.25		GGG	249	0.76
	AGC	88	0.42	组氨酸	CAU	299	1.5
				Histidine	CAC	101	0.5
				谷氨酰胺	CAA	476	1.51
精氨酸	CGU	243	1.36	Glutamine	CAG	153	0.49
Arginine	CGC	100	0.56	半胱氨酸	UGU	150	1.52
	CGA	231	1.29		UGC	48	0.48
	CGG	85	0.48	苯丙氨酸	UUU	646	1.29
	AGA	304	1.7	Phenylalanine	UUC	355	0.71
	AGG	108	0.61	色氨酸			
甲硫氨酸 Methionine	AUG	423	1	Tryptophan	UGG	329	1
亮氨酸	UUA	657	2.03	酪氨酸	UAU	510	1.56
Leucine	UUG	350	1.08	Tyrosine	UAC	142	0.44
	CUU	423	1.31	脯氨酸	CCU	301	1.59
	CUC	133	0.41	Proline	CCC	167	0.88
	CUA	272	0.84		CCA	202	1.07
	CUG	109	0.34		CCG	88	0.46

2.3 大粒香叶绿体基因组长重复序列和 SSR 分析

在大粒香 cpDNA 中检测到十九个长重复序列，包含了八个正向重复，长度范围在 30~52 bp，以及十一个回文重复，其长度范围在 30~127 bp。最长的 127 bp 的重复序列位于 *rps*19-*psb*K 的基因间隔区内，而含有最多长重复序列的区间为 *rac*L-*acc*D。区域位置分布显示绝大多数分布在基因间隔区内(表 4)。

表 4 大粒香 cpDNA 的重复序列

Table 4 The cpDNA repeat sequence of Dalixiang

序号 No.	大小（bp) Size	类型 Type	位置 Position	区域 Location
1	52	正向	56 044	IGS(*rbc*L-*acc*D)
2	35	正向	130 388	IGS(*ndh*B-*trn*L-CAA)
3	32	正向	55 946	IGS(*rac*L-*acc*D)
4	43	正向	12 917	IGS(*trn*G-GCC-*trn*T-GGU)

5	46	正向	63 122	IGS(*pet*A-*pet*G)
6	38	正向	56 005	IGS(*rbc*L-*acc*D)
7	35	正向	84 954	IGS(*trn*H-GUG-*trn*V-GAC)
8	30	正向	86 451	IGS(*trn*H-GUG-*trn*V-GAC)
9	35	回文	85 004	IGS(*trn*H-GUG-*trn*V-GAC)
10	35	回文	84 954	IGS(*trn*H-GUG-*trn*V-GAC)
11	30	回文	128 946	Intron(*ndh*B)
12	46	回文	63 122	IGS(*pet*A-*pet*G)
13	30	回文	86 451	IGS(*trn*H-GUG-*trn*V-GAC)
14	32	回文	55 946	IGS(*rac*L-*acc*D)
15	38	回文	56 005	IGS(*rac*L-*acc*D)
16	52	回文	56 044	IGS(*rac*L-*acc*D)
17	127	回文	0	IGS(*rps*19-*psb*K)
18	30	回文	66 243	CDS(*rps*18)
19	32	回文	57 152	IGS(*acc*D-*psa*I)

注：**LSC**. 大单拷贝区；**IGS**. 基因间隔区；**CDS**. 基因编码区。

Note: **LSC**. Large single copy; **GIS**. Gene intergenic spacer; **GCR**. Gene coding region.

在大粒香 cpDNA 的 129 个 SSR 位点中有 95 个单核苷酸重复,并且 79.07 %的 SSR 由 A 或 T 组成，表明 SSR 位点有使用 A/T 碱基的偏好。同时研究表明 SSR 位点在大粒香 cpDNA 上分布不均，在 LSC 区、SSC 区以及 IR 区分别分布了 95、18 和 16 个 SSR 位点(表 5)。

表 5 大粒香 cpDNA 中的 SSR

Table 5 Distribution of SSR in the Dalixiang cpDNA

序号 No.	重复类型 Type	简单重复序列 SSR	大小 Size(bp)	位置 Location	序号 No.	重复类型 Type	简单重复序列 SSR	大小 Size(bp)	位置 Location
1	c	(AA)$_5$(A)$_{11}$	11	LSC	44	c	(TT)$_4$(T)$_8$	8	LSC
2	c	(A)$_8$(AA)$_4$	8	LSC	45	c	(AA)$_4$(A)$_8$	8	LSC
3	p2	(AG)$_5$	10	LSC	46	c	(A)$_9$(AA)$_4$	8	LSC
4	c	(A)$_{10}$(AA)$_5$	10	LSC	47	c	(TT)$_4$(T)$_8$	8	LSC
5	p4	(TAAA)$_4$	16	LSC	48	c	(T)$_8$(TT)$_4$	8	LSC
6	c	(A)$_8$(AA)$_4$	8	LSC	49	p2	(TA)$_4$	8	LSC
7	c	(A)$_8$(AA)$_4$	8	LSC	50	p2	(AT)$_4$	8	LSC
8	c	(A)$_9$(AA)$_4$	8	LSC	51	c	(TT)$_5$(T)$_{10}$	10	LSC
9	c	(AA)$_4$(A)$_8$	8	LSC	52	c	(AA)$_4$(A)$_9$	9	LSC
10	c	(TT)$_4$(T)$_8$	8	LSC	53	c	(AA)$_4$(A)$_8$	8	LSC
11	c	(A)$_8$(AA)$_4$	8	LSC	54	p2	(AG)$_4$	8	LSC
12	p2	(AG)$_4$	8	LSC	55	c	(TTTT)$_3$(TTT)$_4$(TT)$_6$(T)$_{12}$	12	LSC
13	p2	(GA)$_4$	8	LSC	56	c	(C)$_8$(CC)$_4$	8	LSC
14	c	(TT)$_4$(T)$_9$	9	LSC	57	p4	(GTAG)$_4$	16	LSC
15	p4	(CTTT)$_3$	12	LSC	58	c	(AA)$_4$(A)$_8$	8	LSC
16	c	(AA)$_4$(A)$_8$	8	LSC	59	c	(TT)$_4$(T)$_8$	8	LSC
17	c	(TT)4(T)8	8	LSC	60	p4	(AATA)$_3$	12	LSC
18	c	(AA)$_4$(A)$_8$	8	LSC	61	c	(T)$_9$(TT)$_4$	8	LSC
19	p6	(TTTCTA)$_3$	18	LSC	62	c	(TT)$_4$(T)$_9$	9	LSC
20	p6	(ATAGAA)$_3$	18	LSC	63	c	(A)$_9$(AA)$_4$	8	LSC
21	c	(AA)$_4$(A)$_8$	8	LSC	64	c	(AA)$_4$(A)$_8$	8	LSC

22	c	$(AA)_4(A)_9$	9	LSC	65	c	$(TT)_4(T)_8$	8	LSC
23	p2	$(AT)_5$	10	LSC	66	c	$(AA)_4(A)_8$	8	LSC
24	p2	$(AT)_4$	8	LSC	67	c	$(AA)_4(A)_8$	8	LSC
25	c	$(TT)_4(T)_8$	8	LSC	68	c	$(TT)_4(T)_8$	8	LSC
26	c	$(A)_8(AA)_4$	8	LSC	69	c	$(AA)_4(A)_8$	8	LSC
27	c	$(AA)_4(A)_8$	8	LSC	70	c	$(AA)_4(A)_8$	8	LSC
28	c	$(AA)_4(A)_8$	8	LSC	71	c	$(A)_8(AA)_4$	8	LSC
29	c	$(A)_{11}(AA)_5$	10	LSC	72	c	$(TT)_4(T)_8$	8	LSC
30	c	$(A)_8(AA)_4$	8	LSC	73	c	$(T)_{10}(TT)_5$	10	LSC
31	p2	$(AT)_4$	8	LSC	74	p2	$(TA)_4$	8	LSC
32	p2	$(AT)_4$	8	LSC	75	p2	$(TA)_4$	8	LSC
33	c	$(TT)_4(T)_8$	8	LSC	76	p4	$(AGAA)_3$	12	LSC
34	c	$(T)_8(TT)_4$	8	LSC	77	c	$(TT)_4(T)_8$	8	LSC
35	c	$(T)_8(TT)_4$	8	LSC	78	c	$(A)_9(AA)_4$	8	LSC
36	c	$(AA)_4(A)_8$	8	LSC	79	p4	$(TTTA)_3$	12	LSC
37	c	$(AA)_4(A)_8$	8	LSC	80	c	$(AA)_4(A)_8$	8	LSC
38	c	$(A)_9(AA)_4$	8	LSC	81	c	$(T)_{10}(TT)_5$	10	LSC
39	c	$(TT)_4(T)_9$	9	LSC	82	c	$(A)_{10}(AA)_5$	10	LSC
40	p2	$(CT)_5$	10	LSC	83	c	$(T)_{10}(TT)_5$	10	LSC
41	c	$(TT)_4(T)_8$	8	LSC	84	c	$(T)_8(TT)_4$	8	LSC
42	c	$(TT)_4(T)_9$	9	LSC	85	c	$(T)_8(TT)_4$	8	IR_B
43	c	$(T)_8(TT)_4$	8	LSC	86	c	$(T)_9(TT)_4$	8	LSC
87	c	$(T)_9(TT)_4$	8	LSC	109	p3	$(TAT)_4$	12	SSC
88	c	$(T)_8(TT)_4$	8	LSC	110	c	$(T)_9(TT)_4$	8	SSC

89	c	$(TT)_4(T)_8$	8	LSC
90	p3	$(TCT)_4$	12	LSC
91	c	$(T)_{11}(TT)_5$	10	LSC
92	c	$(T)_8(TT)_4$	8	LSC
93	c	$(A)_8(AA)_4$	8	LSC
94	p2	$(AG)_4$	8	LSC
95	c	$(TT)_5(T)_{10}$	10	LSC
96	c	$(TT)_4(T)_8$	8	SSC
97	c	$(A)_8(AA)_4$	8	SSC
98	c	$(TT)_4(T)_9$	9	SSC
99	c	$(A)_8(AA)_4$	8	SSC
100	c	$(A)_8(AA)_4$	8	SSC
101	c	$(A)_8(AA)_4$	8	SSC
102	c	$(A)_8(AA)_4$	8	SSC
103	c	$(AA)_4(A)_8$	8	SSC
104	p4	$(AACA)_3$	12	SSC
105	c	$(A)_8(AA)_4$	8	SSC
106	p4	$(AATA)_3$	12	SSC
107	c	$(CC)_4(C)_8$	8	SSC
108	c	$(A)_8(AA)_4$	8	SSC
111	p2	$(TC)_5$	10	SSC
112	c	$(A)_8(AA)_4$	8	SSC
113	c	$(AA)_4(A)_8$	8	SSC
114	c	(CC)4(C)9	9	IR$_A$
115	c	$(G)_8(GG)_4$	8	IR$_A$
116	p2	$(CT)_4$	8	IR$_A$
117	p2	$(GA)_4$	8	IR$_A$
118	p4	$(AACG)_3$	12	IR$_A$
119	c	$(AA)_4(A)_8$	8	IR$_A$
120	c	$(TTTT)_4(T)_{17}(TTTTT)_3(TTT)_5(TT)_8$	16	LSC
121	p4	$(TCGT)_3$	12	IR$_B$
122	p2	$(TC)_4$	8	IR$_B$
123	p2	$(AG)_4$	8	IR$_B$
124	c	$(C)_8(CC)_4$	8	IR$_B$
125	c	$(G)_9(GG)_4$	8	IR$_B$
126	c	$(A)_{10}(AA)_5$	10	IR$_B$
127	p2	$(CT)_4$	8	IR$_B$
128	c	$(TT)_4(T)_8$	8	IR$_B$
129	c	$(AA)_4(A)_8$	8	IR$_B$

注：**c**．单碱基；**p2**．2 碱基单元；**p3**．3 碱基单元；**P4**．4 碱基单元；**P6**．6 碱基单元；**LSC**．大单拷贝序列；**SSC**．小单拷贝序列；**IR$_A$**．正向重复序列；**IRB**．反向重复序列。

Note:**c**. Dinucleotide; **p3**. Trinucleotide; **P4**. Tetranucleotide; **P6**. hexanucleotide; **LSC**. large single copy; **SSC**. small single copy; **IR**. inverted repeats.

2.4 大粒香叶绿体基因组系统发育分析

将大粒香与粳稻、籼稻、野生稻及 2 个外类物种等共 15 个叶绿体基因组序列构建发育树。发育树分析表明，15 个物种可分为 3 类，第一类为 *Bambusa oldhamii*，第二类为 *Sorghum bicolor*，第三类为 13 个稻属物种组成。在第三类群中又可分为 4 个小类群，三种野生稻（*O.* australiensis 300316、*O.* meridionalis、*O.* ruifipogon）各为一类，其余 10 种栽培稻为一类。在栽培稻类群中，粳稻与籼稻分别处于不同进化分支。其次，大粒香水稻与粳稻 Tropical Japonica 在同一分支，表明两者的进化关系比其他水稻品种近（图 2）。

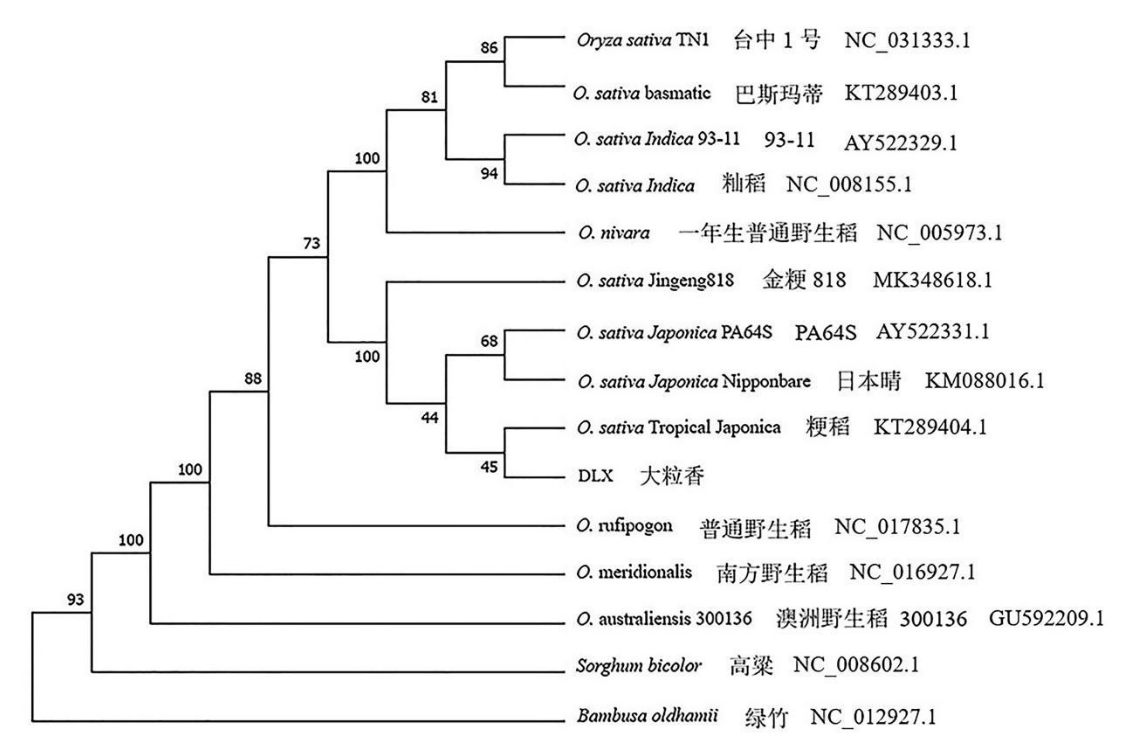


图 2 15 种植物的叶绿体基因组序列的系统进化树

Fig. 2 Phylogenetic tree constructed using chloroplast genome of 15 plants

3 讨论与结论

本研究测得大粒香叶绿体基因组全长为 134 563 bp，GC 含量为 39 %，LSC 为 80 864 bp，SSC 为 12 347 bp，IR 为 20 676 bp，并注释到 129 个基因，与已报道的禾本科数据相符（李裕华等，2020）。前人通过比对不同禾本科植物叶绿体基因组序列表明，叶绿体基因保守程度较高，但一些基因在进化过程中仍然出现退化缺失现象（唐萍等，2011；付涛等，2016）。本研究将大粒香叶绿体基因组与 Wang 等（2016）报道的热带粳稻叶绿体基因组进行比对，结果显示二者在细胞色素 b/f 复合体相关基因、光系统 Ⅰ、Ⅱ 相关基因、核糖体蛋白大和小亚基相关基因、tRNA 和未知功能基因等基因差异较少，但在大粒香叶绿体基因中不存在 *lhb*A 基因。*lhb*A 基因是和光合作用过程中光系统Ⅱ有关的基因，在热带粳稻中存在 *lhb*A 基因，并且在喜好温暖的禾本科植物毛竹中同样也存在 *lhb*A 基因（Xin et al., 2016），这可能是热带粳稻为适应热带光温生长环境中逐渐进化而得。

SSR 位点可被用于辅助育种和遗传连锁作图等方面的研究，而 cpDNA 具有序列保守、结构稳定、易测序等优点，有助于解决类群间的遗传多样性（Powell et al., 1995; Pugh et al.,

2004;Song et al., 2019）。首先,本研究结果表明大粒香 cpDNA 的 SSR 位点对 A/U（T）碱基有着明显偏好，这与前人研究结论相一致（张慧等，2018；李泳潭等，2020；郑祎等，2020；王一麾等，2021；吴朝昕等，2021）。其次，大粒香 cpDNA 的密码子也偏好 A/U（T）碱基，这种密码子使用情况也存在于其他物种中（Shinozaki et al., 1986;Ohyama et al., 1988;郑祎等，2020；朱斌等，2021；吴朝昕等，2021）。大粒香 cpDNA 编码蛋白质的密码子和 SSR 位点都偏好 A 碱基或者 U（T）碱基，可能是造成大粒香 cpDNA 总 A/U（T）含量大于总 GC 含量的原因。根据 Niu 等（2007）的研究报道,A/T 核苷酸含有 7 个氮原子，比 G/C 核苷酸少一个，所以富含 A/T 核苷酸耗能更少，有利于 cpDNA 的复制。这也可能是大粒香 cpDNA 富含 A/U（T）的原因。

系统发育基因组学是利用分子数据研究生物间发育关系。cpDNA 的具有序列保守，结构稳定，易测序等优点，因此基于 cpDNA 进行的系统发育研究得到很好的发展（Eisen.,1998; Eisen &Hanawalt,1999; Delsuc et al., 2005）。本研究对优质稻大粒香叶绿体基因组测序数据进行系统发育分析，研究表明大粒香与热带粳稻聚为一类；粳稻与籼稻不为一类，这一结果与林张翔等（2014）的研究结果相同，支持了 Huang 等（2012）的籼粳稻起源假说。

综上所述，本研究获得的大粒香的叶绿体基因组大小、结构、基因数量、重复序列、密码子偏好、系统发育树等特征信息，将为进一步研究大粒香的系统进化和育种研究提供理论依据。

参考文献

DELSUC F, BRINKMANN H, PHILIPPE H, 2005. Phylogenomics and the reconstruction of the tree of life[J]. Nat Rev Genet, 6(5): 61-375.

EISEN JA, 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis[J]. Genome Res, 8(3): 163-167.

EISEN JA, HANAWALT PC, 1999. A phylogenomic study of DNA repair genes, proteins, and processes[J]. Mutat Res-DNA Repair, 435(3): 171-213.

FAN J, ZHU WY, LI ZF, et al., 2020. Chloroplast genome sequence of a yellow colored rice (*Oryza sativa* L.): insight into the genome structure and phylogeny[J]. Mitochondrial DNA Part B, 5(3): 3650–3652.

FANG L, YAN Z, DENGJIE L, et al., 2017. The complete chloroplast genome sequence of *Oryza rhizomatis*(Poaceae)[J]. Mitochondrial DNA Part B, 2(2):467–468.

FU T, WANG ZL, QIAN PX, et al., 2016. The latest research progress and application of the DNA barcode in higher plants[J]. J Nucl Agric Sci, 30(5): 887-896. [付涛，王志龙，钱萍仙，等，2016．高等植物 DNA 条形码最新研究进展及其应用[J]．核农学报，30(5)：887-896.]

HUANG XH, KURATA N, WEI XH, et al., 2012. A map of rice genome variation reveals the origin of cultivated rice[J]. Nature, 490(7421): 497–501.

JEON JH, KIM SC, 2019. Comparative analysis of the complete chloroplast genome sequences of three closely related east-Asian wild roses（*Rosa sect*. synstylae Rosaceae）[J]. Genes, 10 (1): 1-14.

JIANG ZQ, 2008. Breeding and application of Dalixiang, a new quality rice line[J]. Guizhou Agric Sci, 36(5): 12-13.[蒋志谦，2008．优质水稻新品系大粒香的选育及应用[J]．贵州农业科学．36(5)： 12-13.]

LI YH, REN YK, ZHAO XH, et al., 2020. Research Progress on Chloroplast Genome of Major Gramineous Crops[J]. BiotechnolBull, 36(11):112-121. [李裕华，任永康，赵兴华，等，2020.

禾本科主要农作物叶绿体基因组研究进展[J]．生物技术通报，36(11)：112-121.]

LI XY, XIAO BG, GAO YL, et al., 2011. Analysis of SSR loci in chloroplast and mitochondrial genomes of tobacco[J]. Acta Bot Boreali-OccidentaliaSinica, 31(12): 2399-2405.[李绪英，肖炳光，高玉龙，等，2011．烟草叶绿体基因组和线粒体基因组 SSR 位点分析[J]．西北植物学报，31(12)：2399-2405.]

LI YT, ZHANG J, LI LF, et al., 2018. Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis*—“wild plants with a tiny population”—and three other *Pyrus* species[J]. Int J Mol Sci, 19(3262):1-19.

LUO RF, LUO J, MEI YX, et al., 2012. The purification, rejuvenation and application of good quality rice Dalixiang of maogong brand[J]. Seed, 31(9): 131-132.[罗仁发，罗节，梅映雪，等，2012．茅贡牌优质稻大粒香提纯复壮及应用[J]．种子，31(9)：131-132.]

LIN ZX, WANG YY, FU F, et al., 2014. Complete chloroplast genome of Dongxiang wild rice and its application in phylogenetic analysis[J]. J of Zhejiang Univ (Agric & Life Sci),40(4): 397-403.[林张翔，王营营，付菲，等，2014．东乡野生稻叶绿体基因组拼接及系统进化分析[J]．浙江大学学报(农业与生命科学版)，40(4)：397-403.]

LI YT, ZHANG J, HUANG YL, et al., 2020. Analysis of chloroplast genome of *Pyrus betulaefolia*[J]. Acta HorticulturaeSinica, 47(6):1021-1032.[李泳潭，张军，黄亚丽，等，2020．杜梨叶绿体基因组分析[J]．园艺学报，47(6)：1021-1032.]

NIU ZT, XUE QY, WANG H, et al., 2017. Mutational Biases and GC-Biased Gene Conversion Affect GC Content in the Plastomes of *Dendrobium* Genus[J]. Int J Mol Sci, 18(11): 2307-2321.

OHYAMA K, FUKUZAWA H, KOHCHI T, et al., 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome: I. Cloning and gene identification[J]. J Mol Biol, 203(2): 281–298.

POWELL W, MORGANTE M, MCDEVITT R, et al., 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines[J]. Proceed Natl Acad Sci, 92(17): 7759-7763.

PUGH T, FOUET O, RISTERUCCI AM, et al., 2004. A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers[J]. Theor Appl Genet, 108(6): 1151-1161.

SONG Y, CHEN Y, LV JZ, et al., 2019. Comparative chloroplast genomes of *sorghum* species: Sequence divergence and phylogenetic relationships[J]. Biomed Res Int,11(5046958): 1-11.

SHINOZAKI K, OHME M, TANAKA M, et al., 1986 . The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression[J]. Embo J, 5(9):2043-2049.

TANG P, RUAN QY, PENG C, 2011. Phylogeny in structure alterations of *poaceae*cpDNA[J]. Chin Agric Sci Bull, 27(30): 171-176. [唐萍，阮秋燕，彭程，2011．禾本科植物叶绿体基因组结构的系统进化研究[J]．中国农学通报，27(30)：171-176.]

WANG S, GAO LZ, 2016. Complete chloroplast genome sequence and annotation of the tropical japonica group of asian cultivated rice (*Oryza sativa* L.)[J]. Genome Ann, 4(1):1-2.

WANG YH, XIE YF, ZHANG ZX, et al., 2021, The complete chloroplast genome of *Sloanea* sinensis and the systematic status of *Elaeocarpaceae*[J]. Guihaia, 42(1):39-48. [王一麾，谢宜飞，张志翔，等．猴欢喜叶绿体全基因组及杜英科系统地位分析[J]．广西植物 42（1）：39-48.]

WU CX, XU HF, LIU XW, et al., 2021. Analysis of Chloroplast Genome of ‘Goudang 3’[J/OL]. Mol Plant Breeding,1-29[2022-03-30]. http://kns.cnki.net/kcms/detail/46.1068.S.20220118.1548.006.Html.[吴朝昕，徐海峰，刘雪薇，等，2021．‘苟当 3 号’水稻叶绿体基因组特征分析[J/OL].分子植物育种:1-29[2022-03-30].http://kns.cnki.net/kcms/detail/46.1068.S.20220118.1548.006.html.]

XIN Y, YUN-HONG T, YING-YING L, et al., 2016. Chloroplast genome structure in *Ilex* (*Aquifoliaceae*)[J]. Sci Rep,6(1):1-10.

ZHANG X, RONG C, QIN L, et al., 2018. Complete chloroplast genome sequence of *Malus hupehensis*: genome structure, comparative analysis, and phylogenetic relationships[J]. Mol, 23 (2917):1-17.

ZHANG H, HE SB, KONG FD, et al., 2018. Sequence of chloroplast genome and the phyletic evolution within *Leonurus*[J]. Inf Trad Chin Med, 35(4): 21-27.[张慧，何帅兵，孔繁德，等，2018．益母草叶绿体基因组序列与系统进化位置分析[J]．中医药信息，35(4)：21-27.]

ZHEN W, ZHANG H, WANG QM, et al., 2020. Complete chloroplast genome sequence of *Cliviaminiata* and its characteristics[J]. Acta Hortic Sin, 47(12): 2439-2450.[郑祎，张卉，王钦美，等．大花君子兰叶绿体基因组及其特征[J]．园艺学报，47(12)：2439-2450.]

ZHU B, GAN CC, WANG HC, 2021. Characteristics of the complete chloroplast genome of *Dendrobium thyrsiflorum* and Its phylogenetic relationship analysis[J]. Biotechnol Bull, 37(5):38-47.[朱斌，甘晨晨，王洪程，2021．球花石斛(*Dendrobium thyrsiflorum*)叶绿体基因组特征及亲缘关系解析[J]．生物技术通报，37(5)：38-47.]